Simulating Receptor-Flexible Small Molecule binding Using AutoDock

S. Ravichandran, Ph.D.
Advanced Biomedical Computing Center (ABCC)
National Cancer InstituteFrederick
Frederick, MD 21702

05/25/2004
Email: sravi@ncifcrf.gov
Tel: 3018461991
http://nciiris.ncifcrf.gov/~ravichas/docking

Goals of Docking

Fitting a small molecular (drug molecule) into a protein

Docking two proteins together

novel inhibitor discovery using molecular databases

Things U need

1) Structures of protein/small molecule

Molecule	Databases	Method
Protein/DNA	PDB	x-ray, NMR
Small Molecules	CSD	x-ray

2) Software:

Program to do docking: Dock, AutoDock Ludi, FlexX 3D-Dock etc.

3) Powerful Computers: CPU, disk-space etc.

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

?s to ask before docking:

Critical analysis of the protein structures

What to do with bound ions, water molecules? Remove them? Why?

Crystal structure - Rigid nature- bound ions and water are held in the active site. Are they valid configurations?

Resolution of the structure? High resolution/thermal factors

NMR: Which conformation/model to choose?

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

$$E + I \rightarrow [E--I]$$
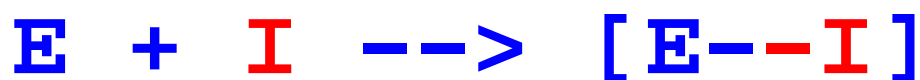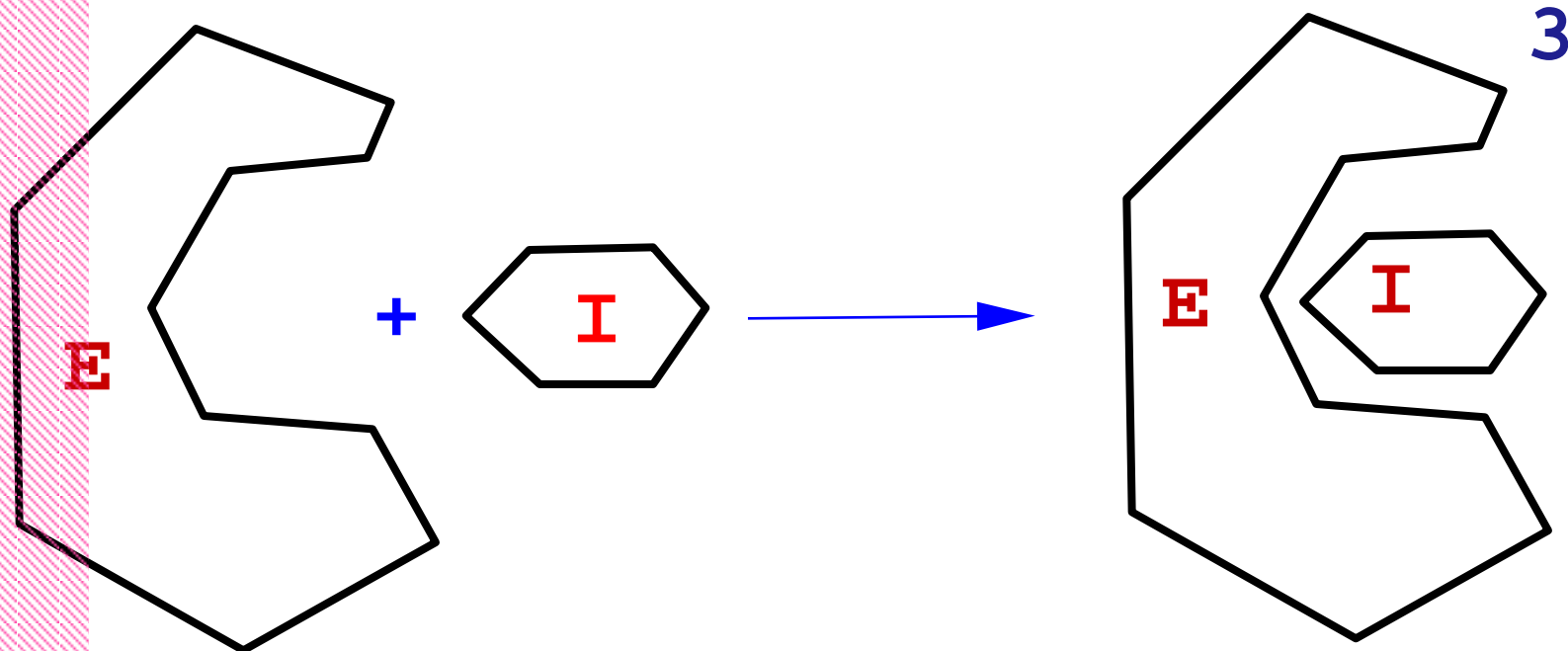
Loosely bound complex

The docking activity gives the proteins the ability to promote and inhibit(or accelerate or prevent) certain chemical reactions.
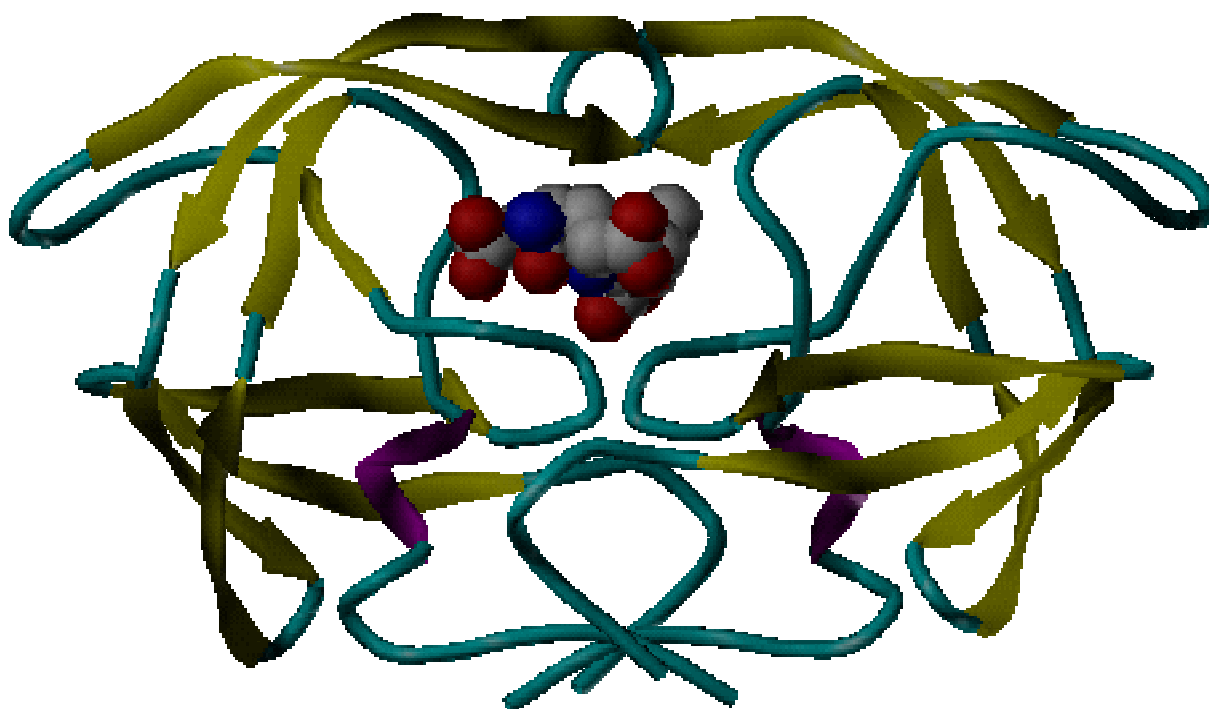
Importance:

Designing bioactive compounds, Computer-Aided Drug design
....
....
...... and many more

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

HIV-Protease complex with tripeptide inhibitor

Things we know:

**a) HIV protease: enzyme in the AIDS virus, imp–
        ortant for its replication**
**b) Chemical reaction takes place in protease
    at an active site**
**c) Inhibitor drugs bind to the active site and
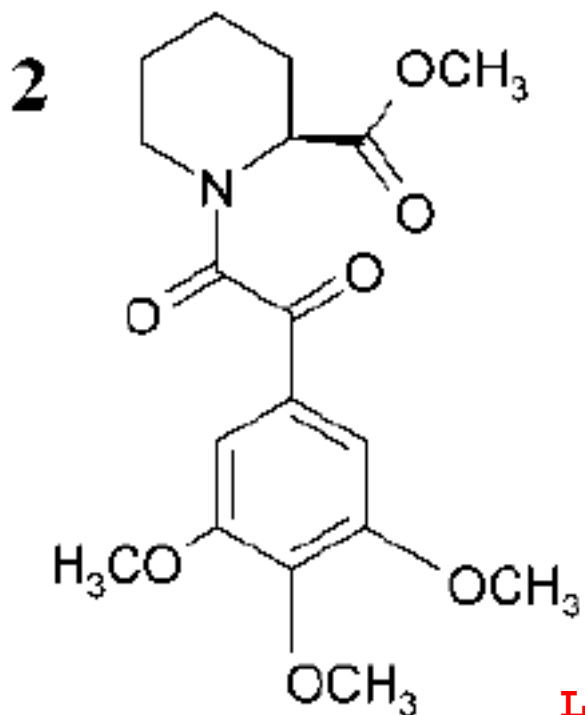   block the functioning**

?s we ask:

a) Binding affinity of the drug
b) What has to be changed for better binding

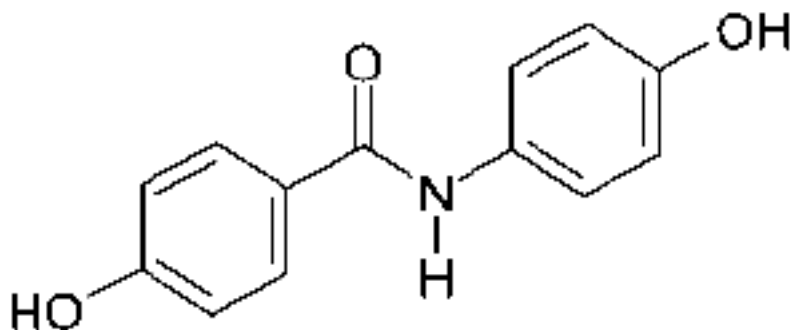S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Drug Design

Immunophilin FKBP soluble receptor for the natural immunosuppressant drug FK506

Aim: Find strong-binding ligands that can substitute for the natural FK506
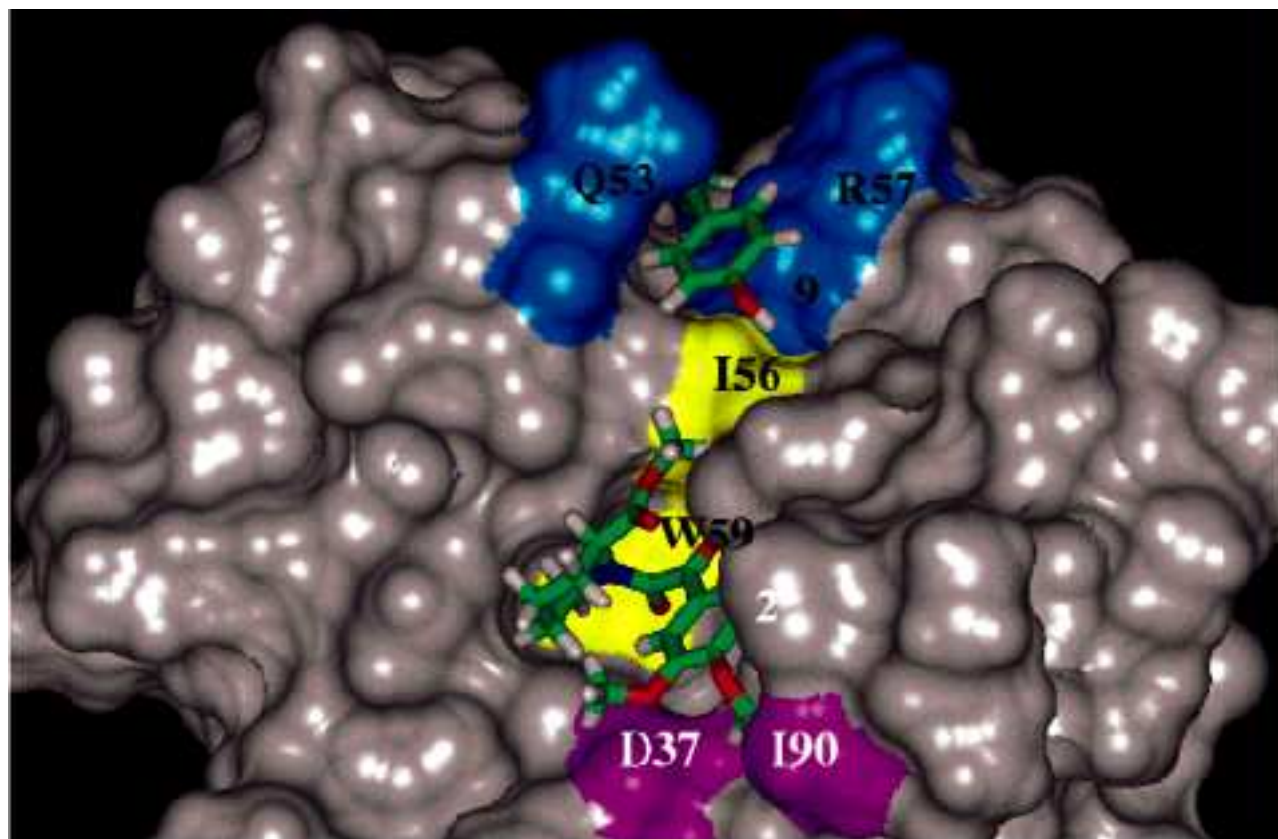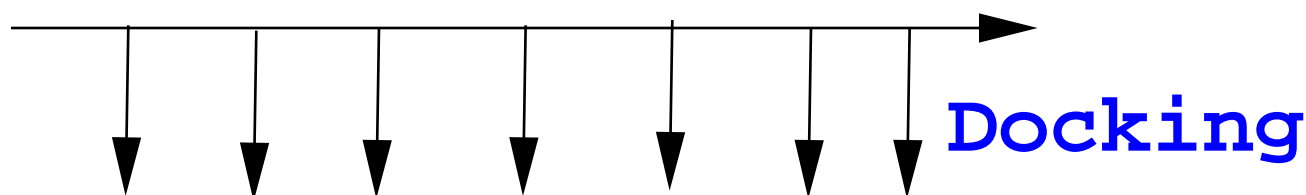

Lin et al, JACS 2002, 124, 5632

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Figure 4. Location of **2** and **9** in the docked complex. **9** was docked in the presence of **2**.

AMBER was used to perform MD & snap-shot conformations were taken and docked using AutoDock

MD

Docking

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Problem:

To find molecular binding sites (hot-spots) by computer
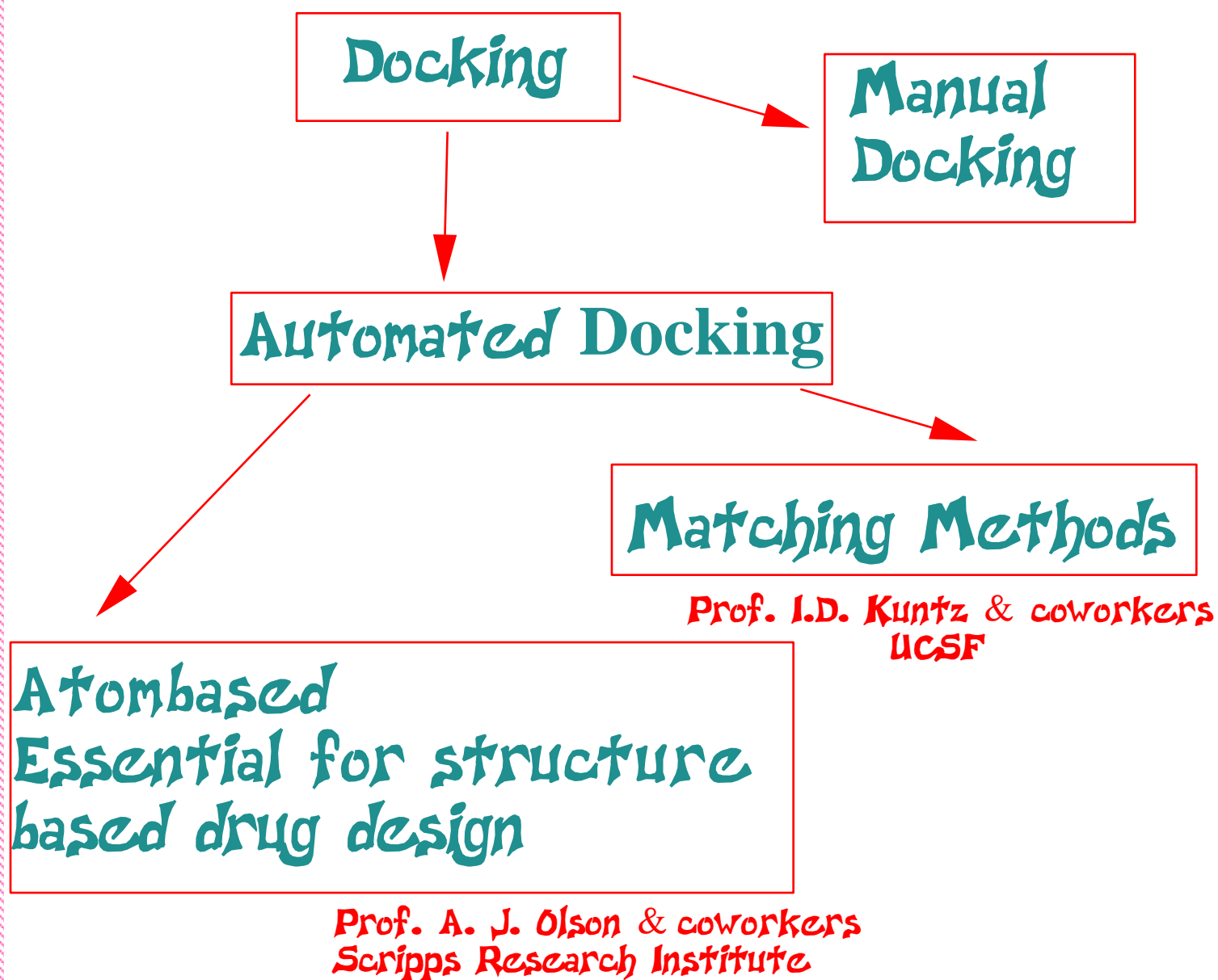
Why is it difficult?

2 Rigid molecules

6 degrees of freedom (3-rotational, 3 translational)

In addition if you assume the small drug sized molecules to be flexible (say 14 degrees of free dom)

1028 variations

Even difficult with a SUPER COMPUTER!!!!

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

Docking → Manual Docking

Docking → Automated Docking

Automated Docking → Matching Methods

Prof. I.D. Kuntz & coworkers UCSF

Automated Docking → Atombased Essential for structure based drug design

Prof. A. J. Olson & coworkers Scripps Research Institute

Atom-based tools are usually slower than the matching methods & considered also better becos of the assumed flexibility of the ligands.

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

AutoDock

1990: Drs. David S. Goodsell;
Arthur J. Olson;
Garrett M. Morris;
Ruth Huey; Scott Halliday
RikBelew...

Scripps Research Institute
La Jolla, CA

Orginal Version: f77
Currently: C++
(ver 3.0.5)

AutoDockTools (ADT)

Free for Academics

Michel Sanner & Ruth Huey
Scripps Resarch Institute

GUI to prepare and submit jobs
for AutoDock
      Code written in Python

- **Optimization**
  Global minimum or maximimum of a function with following properties:

  Continuous function
  Domain Range

Search Techniques:

Local: **Operation which iteratively improves its estimate of minimum by searching for better solutions in a local neighbourhood of current solution**

Global: **Will perform a sophisticated search across several multiple local minimum**
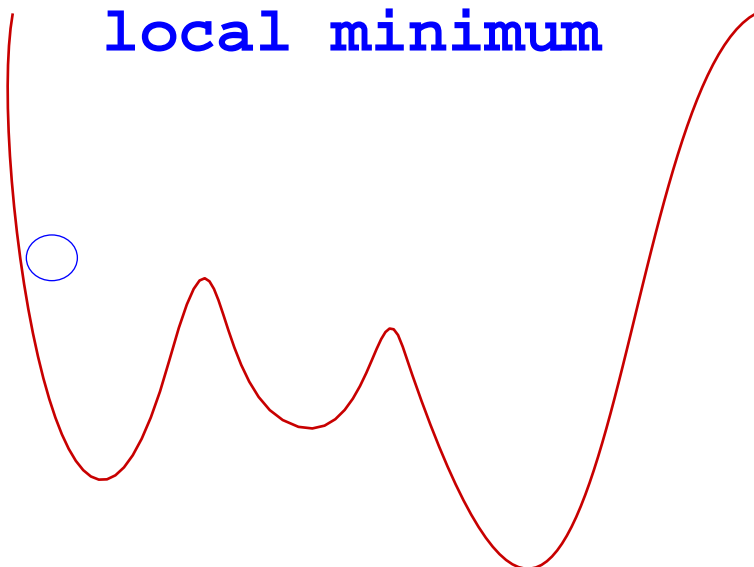
$$\text{Energy} = V_{vdw} + V_{coul} + V_{hb}$$

$$V_{vdw} = \sum_{i,j} 4\varepsilon_{ij} \left[(A_{ij}/r_{ij})^{12} - (B_{ij}/r_{ij})^6 \right]$$

$$V_{coul} = \sum_{i,j} (q_i q_j)/[4\pi\varepsilon(r)\varepsilon_0 r_{ij})$$
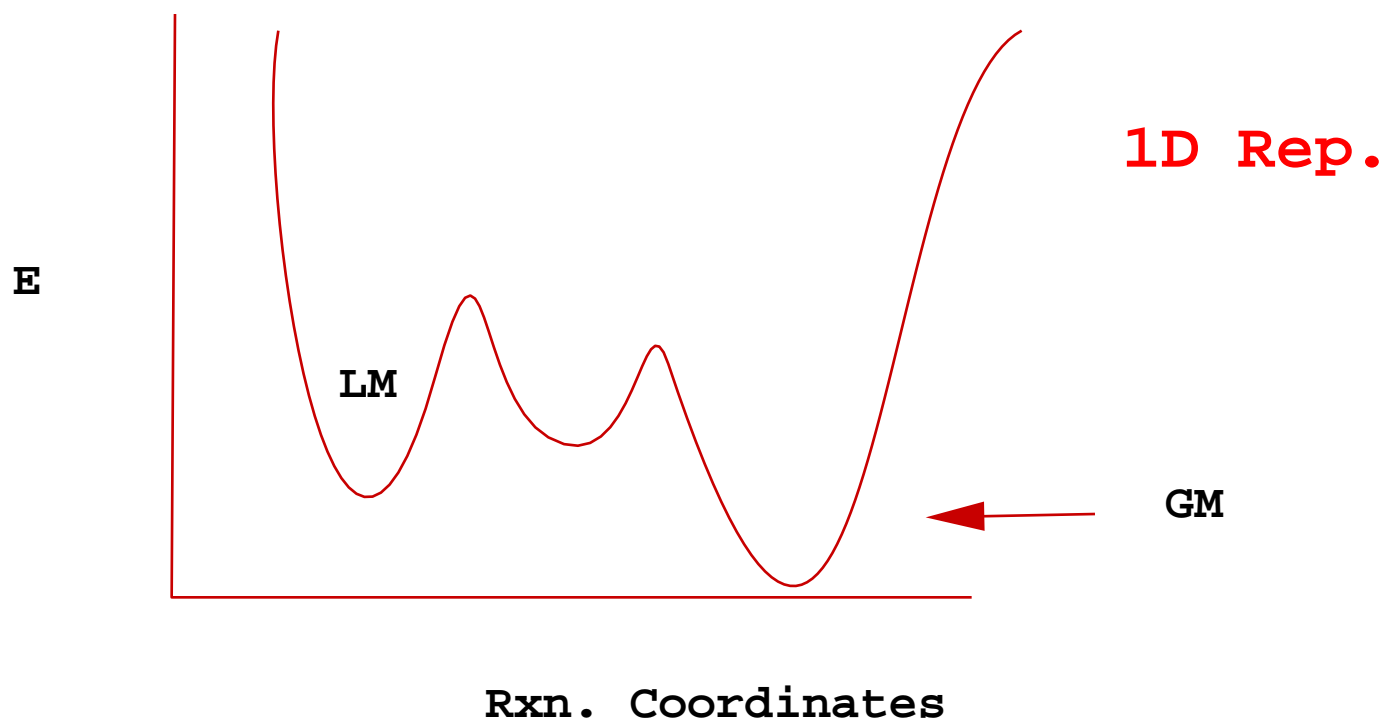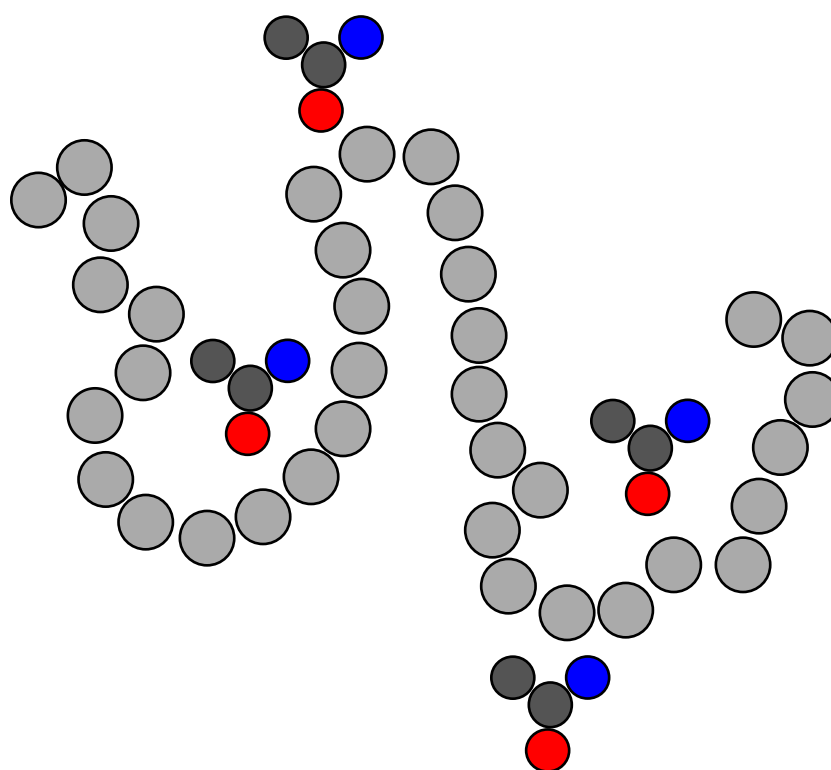
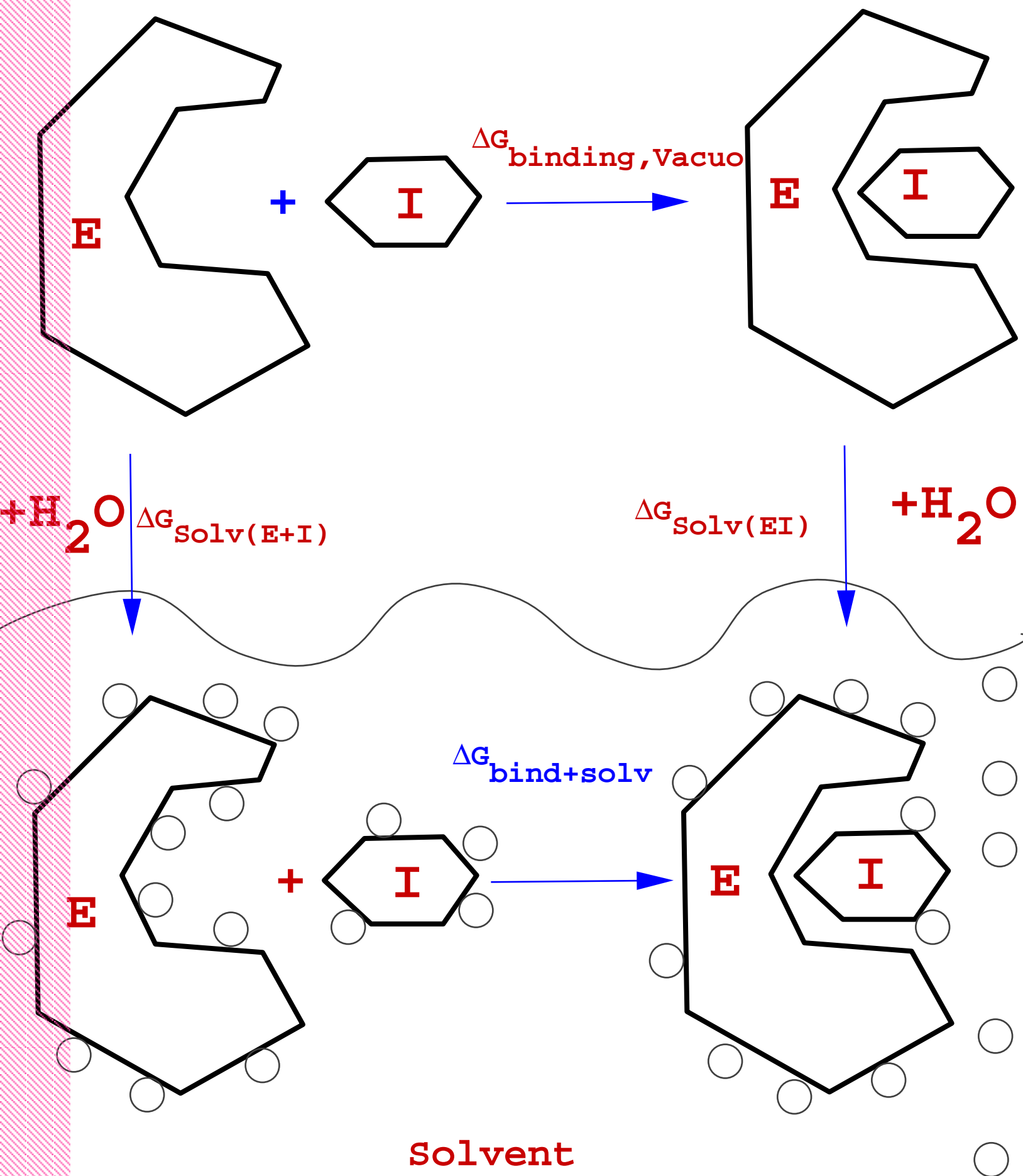**Pairwise
Additive**   (Const dielectric or distance dependent)
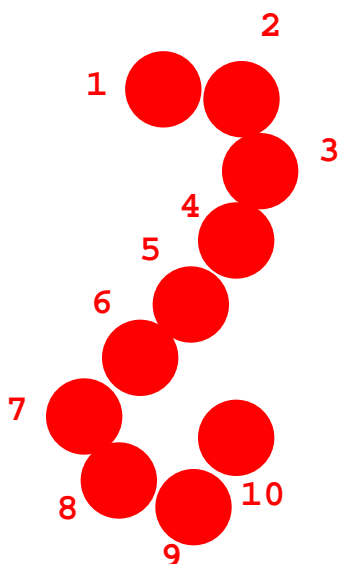
H- bonding 12-10 form is used

Empirical relationship between molecular structure

and binding freee energy

$$\Delta G = K_{vdw} * V_{vdw} + K_{hb} * V_{hb} + K_{ele} * V_{coul}$$
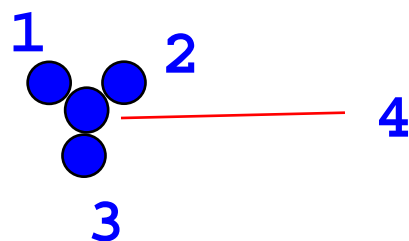$$+ K_{tor} * V_{tor} + K_{sol} * V_{sol}$$

Coefficients are empirically determined using

linear regression analysis from a set of

Protein-ligand complexes with known binding

constants.

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

1D Rep.

E

LM

GM

Rxn. Coordinates

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

$\Delta G_{binding,Vacuo}$

E + I → E I

$+H_2O$ $\Delta G_{Solv(E+I)}$

$\Delta G_{Solv(EI)}$ $+H_2O$

$\Delta G_{bind+solv}$

E + I → E I

Solvent

S.Ravichandran, ABCC
NCI, sravi@ncifcrf.gov

(Representation is based on AutoDock Manual)

Model Receptor (NR = 10)

Model Ligand (NL = 4)

Energy Calculation:

11 12 13 14 15 16 17 18 19 110

21 22 23 24 25 26 27 28 29 210

31 32 33 34 35 36 37 38 39 310

41 42 43 44 45 46 47 48 49 410

Total Terms = NL * NR

To deal at each simulation step

Example: Biotin has 19 atoms
Receptors of the range of
1000 or more atoms

(1000 X 19 = 19000) terms !!!!

AutoDock we will be doing
million steps on GA

so, 19000 * Millon Interactions
per Simulation

?

How to avoid this Computational
Heaviness without losing information?

AutoDock uses GRID(s) to precalculate
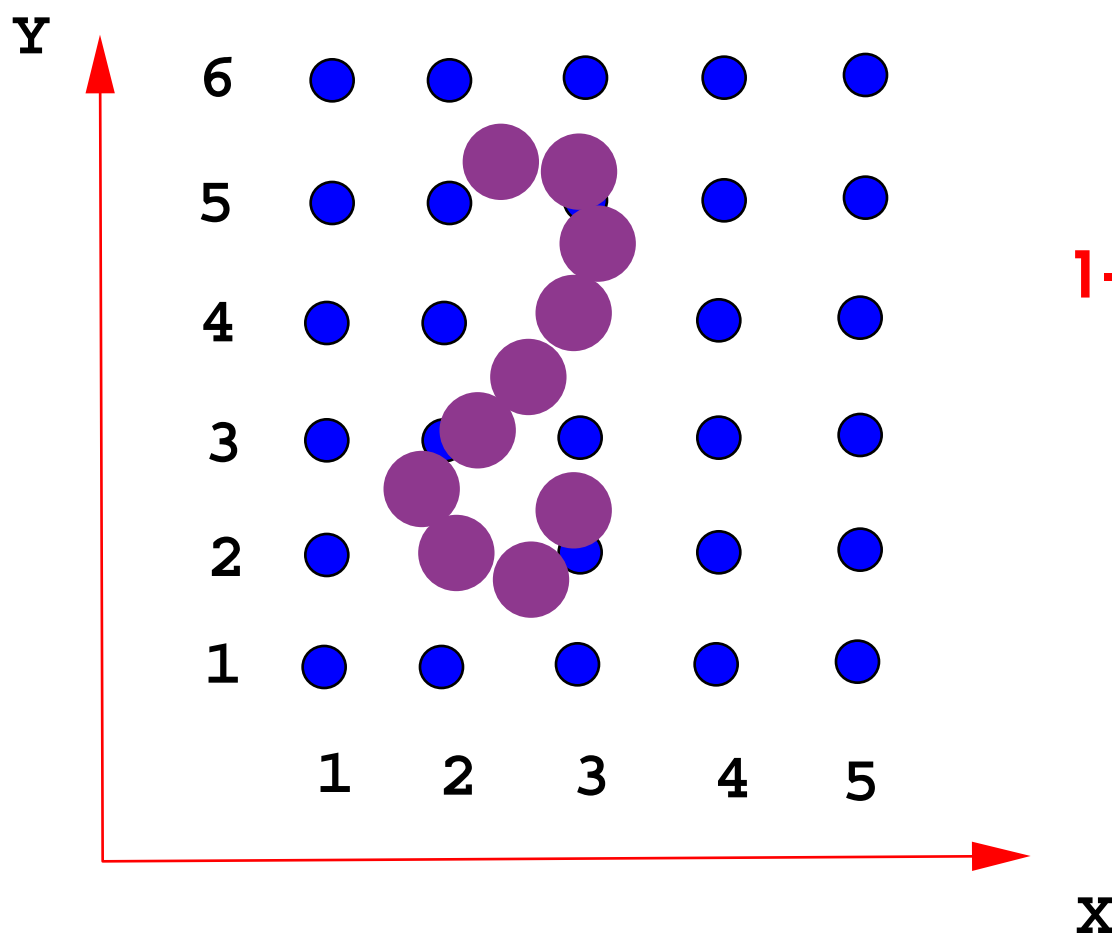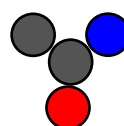the energies due to all the atoms
of the ligands and store them.

These grids are like LOOKUP-TABLES
and used to lookup the interaction
energies later

The Grid Method makes the energy
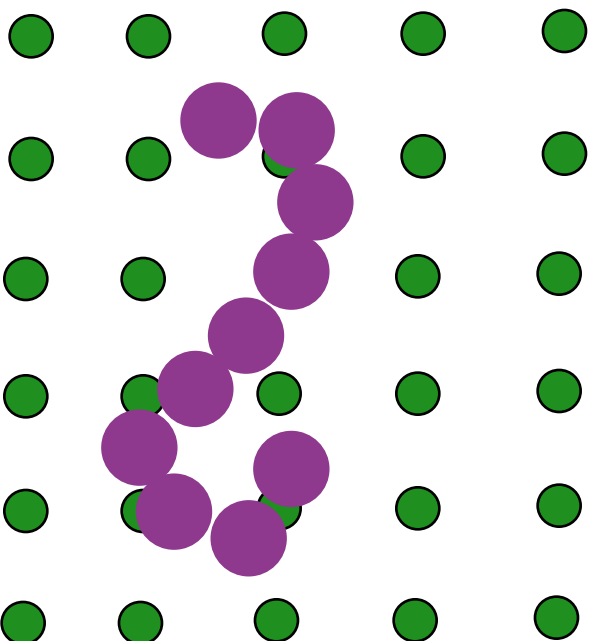calculations independent of the
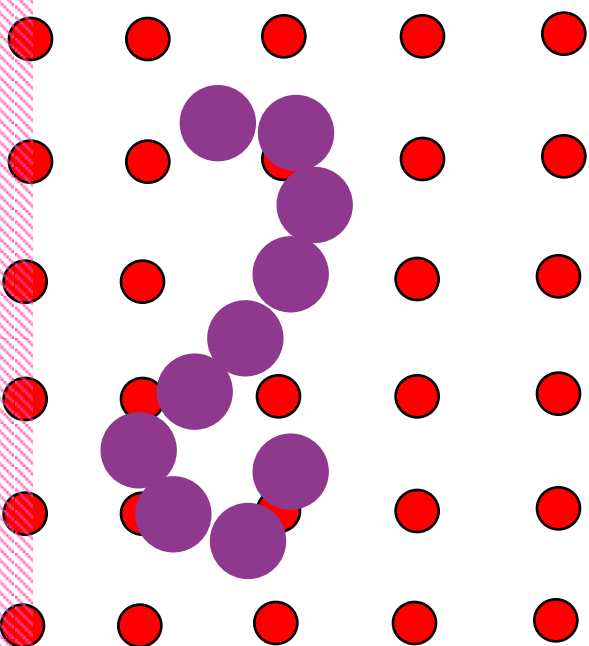size (number) of the Receptor!!

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

STOP

That is all OK, What is a GRID?

Imagine the ligand has 4 atoms of 3 atom types ● ● ●

Y

6

5

4

3

2

1

1 2 3 4 5

X

1-D Rep.
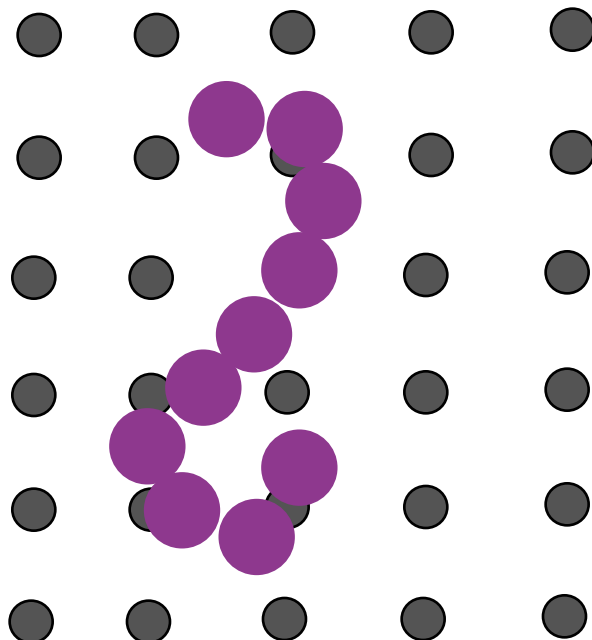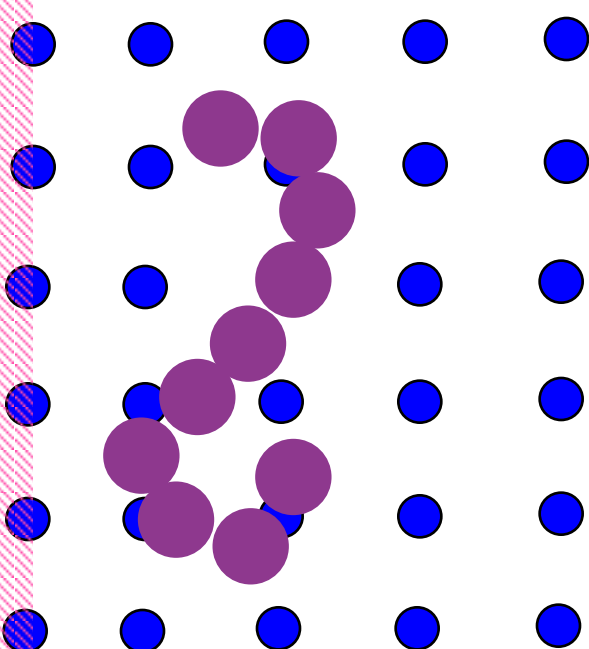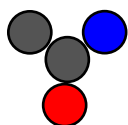
**(or one can do
a PB calculation
to calculate φ)**

**Electron
MAP
If you want
to study
Electrostatics?**

Automated Docking

Conflicting Requirements

a) desire for a robust and physically relevant procedure

b) Computation demands at a reasonable level

Static Receptor – Mobile small drug molecule

Grid–based model

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Methods of docking:
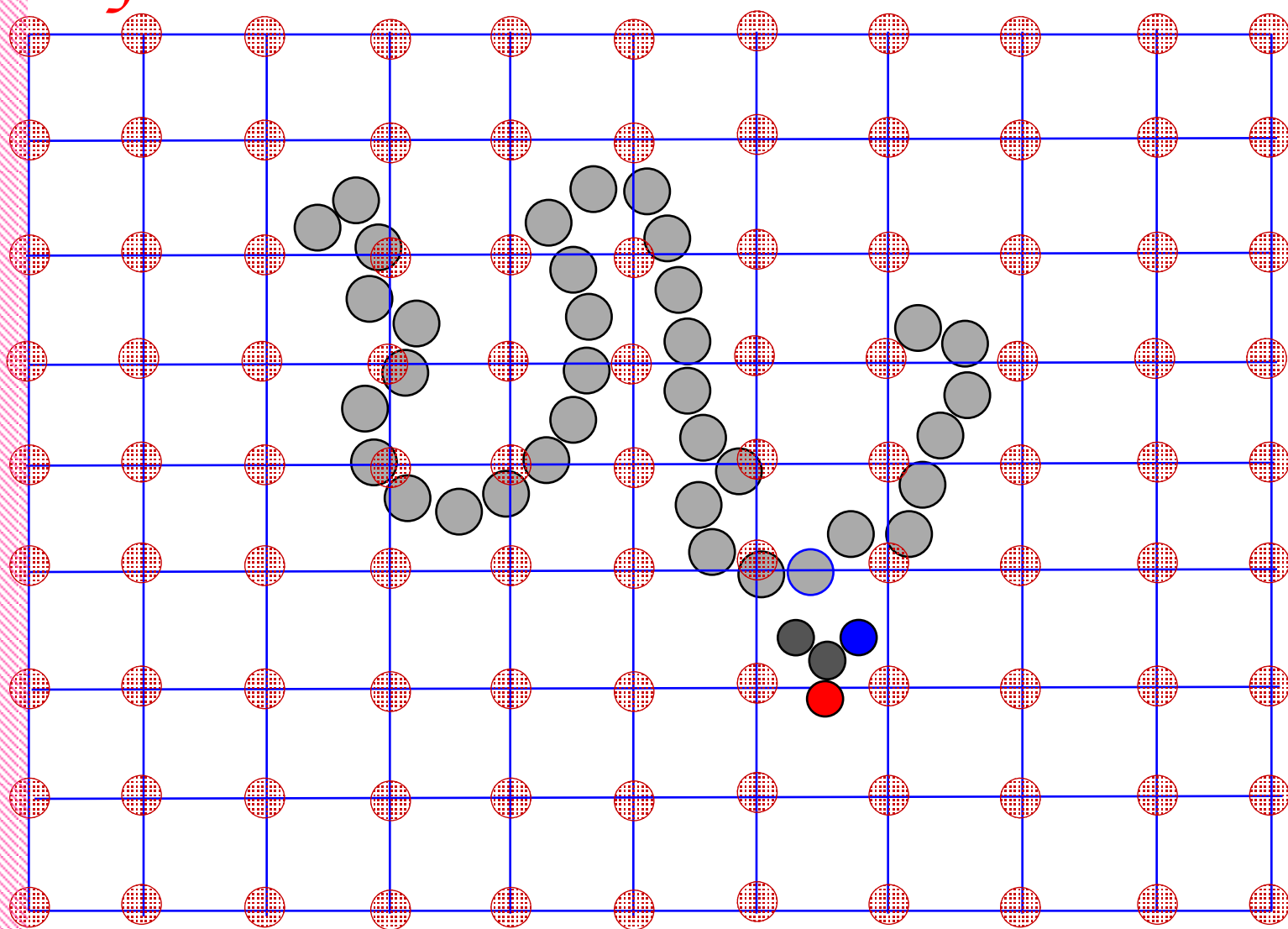
a) Simulated Annealing:
b) Genetic Algorithms:

Annealing is a process in which temperature of the substance is reduced (slowly) until the material crystallizes in a single crystal (usually corresponds to global minimum free–energy)

Simulated Annealing: Computational method of mimicking annealing.
SA can be used to do both global and local search. Global at high Temperature and local at low temperatures.

1-D Representation

E

LM

GM

Rxn. Coordinates
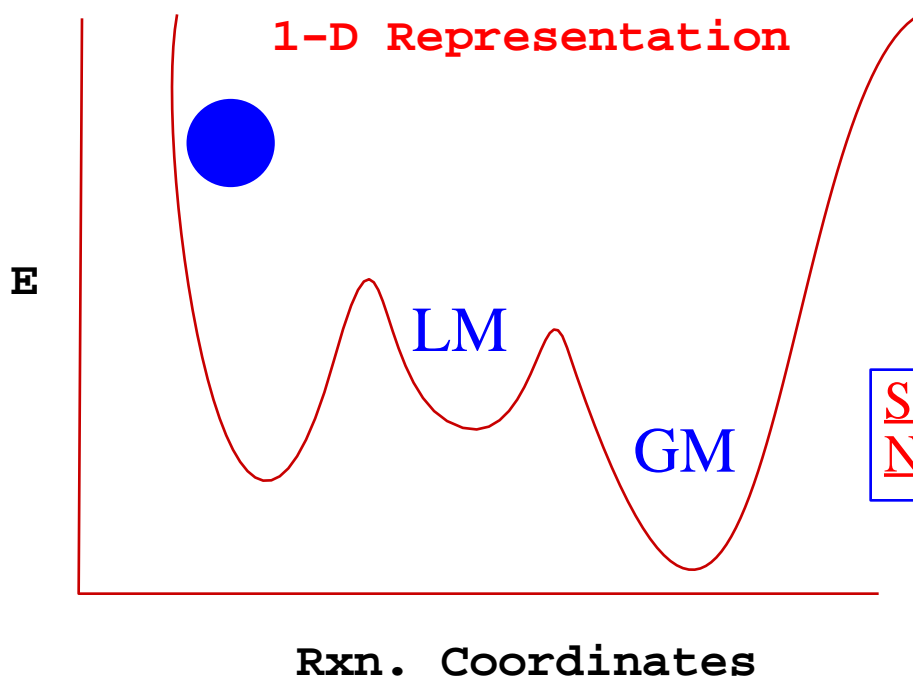
S.Ravichandran, ABCC NCI, sravi@ncifcrf.gov

Simulated Annealing

Monte Carlo

Cycle

Random changes to ligand position, orientation and/or conformation

Compute new E_1

Is E_1 is lower than E_0?

$$T_i = (gT_{i-1})$$

Yes

NO

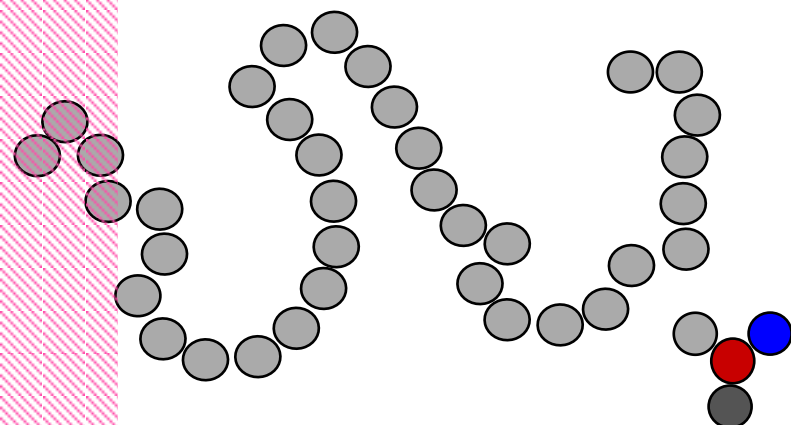Accept the changes

accepted with prob $e^{(-\Delta E/k_B T)}$

Is the number of acceptances or rejection are within the chosen number?

Receptor

No

Yes

End Cycle

Ligand

Does random walk around the receptor

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Evolutionary Computing

I. Rachenberg 1960

Genetic Algorithm

(ideas taken from Natural Genetics & Biological Evolution)

Darwin, John Holland, 1975
John Koza, 1992......

Genetic Programming

Living Organism

CELL

Same set of Chromosomes (strings of DNA)

Serves as a model for the whole organism

Genetic Algorithms are usually used to carry out Global Search.

Genes

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

Hybrid Genetic Algorithm

How big?

Set of solutions (Populations or Chromosomes)

What are chromosomes?

Generate population of n chromosomes

What is fitness?

Evaluate the fitness f(xi)

Crossover how often? Mutation how often? Elitsm? Selection?

Create new population
Elitsm Selection Crossover Mutation

No

Is Termination Condition (Max # of Energy Eval. or Generations) satisfied?

Yes

No Explicit Termination conditions

S.Ravichandran, ABCC NCI, sravi@ncifcrf.gov

STOP

Chromosome 1

E1
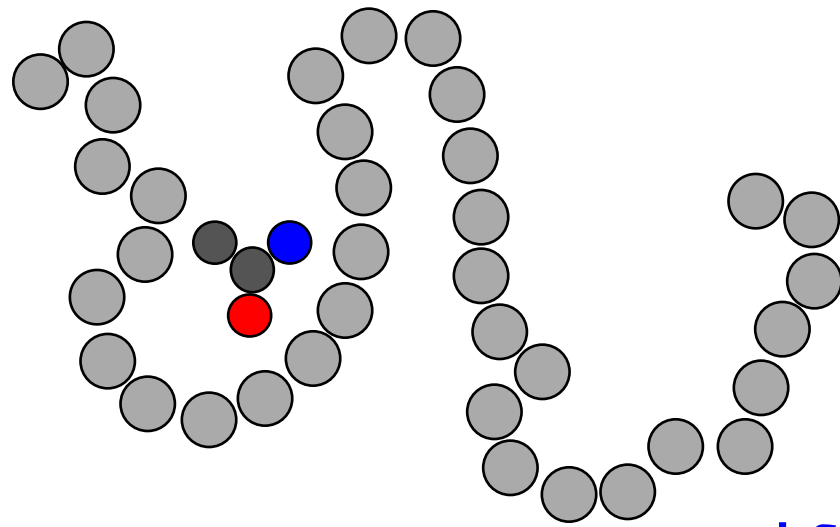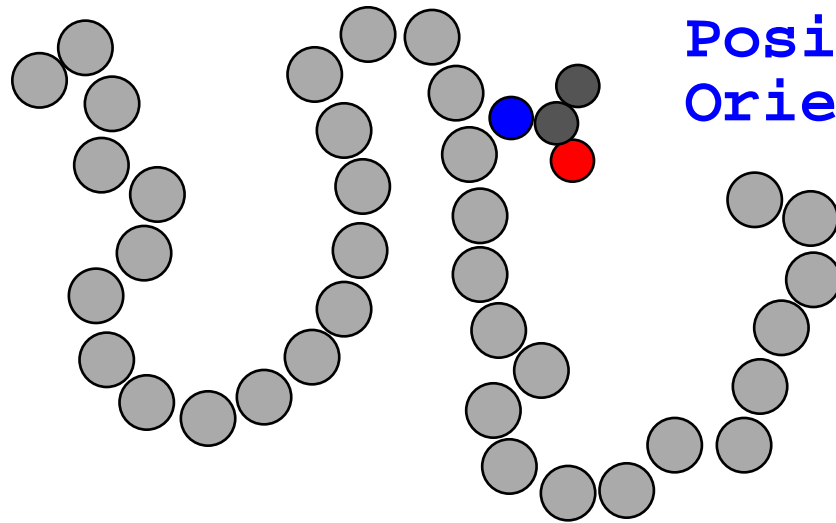
Different
Position &
Orientation

Chromosome 2

E2

and so on .....

If E1 is lower than E2
probably 1 will eventuvally
survive compared to 2

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

x	y	z	q_0	q_1	q_2	q_3	τ_1	τ_2	τ_3	τ_4	τ_5
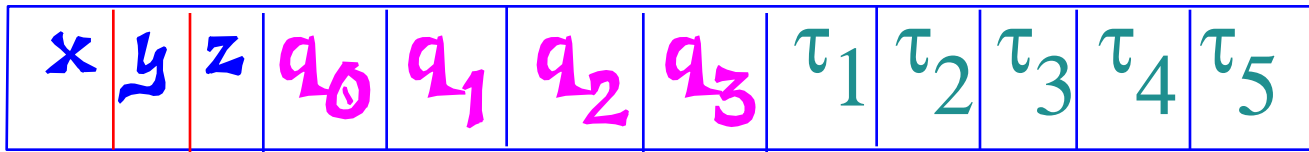
Transl.Gene   **Rot. Gene**   **Torsion Gene**

Generation

Mapping & fitness *translates genotypes–phenotypes to calculate E*

Fitness *Which individuals will reproduce based on worst energy individuval*

Crossover *2–point crossover*

Parent 1

+

Parent 2

=

Offspring

Mutation *random changes to variables using Cauchy distribution*

Elitism *How many top individuvals Survive into next generation?*

S.Ravichandran, ABCC NCI, sravi@ncifcrf.gov
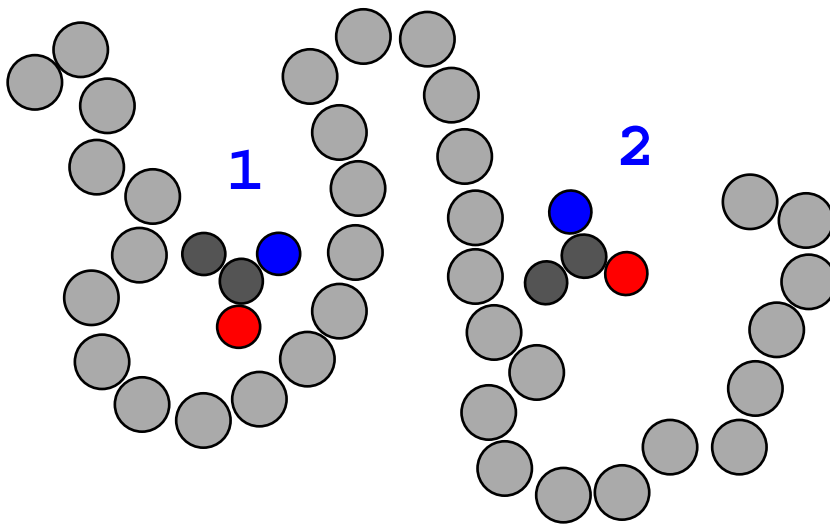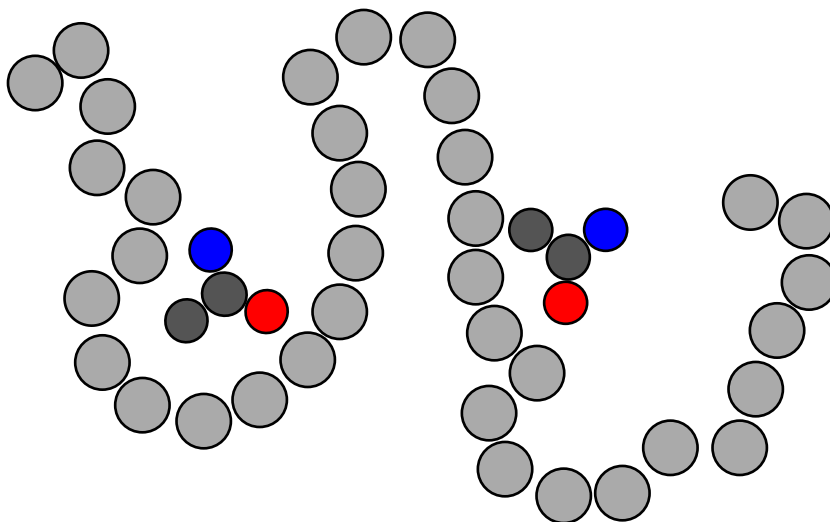
Issues with GA:

1) Premature convergence

2) Too long for convergence
   How many runs?
   How many energy
      evaluations?

3) Choice of parameters

Initialize population

Repeat

*Evaluate Solutions in the popul-
ation*
Perform Competitive Search
Apply Genetic Operators
Perform Local Search

Until Convergence Criteria Satisfied

Pseudo Code for GA

**Ph.D. Thesis of
W.E. Hart (1994)**

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

AutoDock Tools (ADT)

Freeware

Dr. Micheal Sanner
Molecular Graphics Lab
Scripps Research Institute
La Jolla, CA

ADT provides GUI interface for setting up and using AutoDock

`http://www.scripps.edu/pub/olson-web/people/sanner`

1) Sybyl

  Receptor:

  PDB –> Add essential H –> Fix charges Kollman United charges –> save as Mol2 file

Ligand:

PDB –> check for atom types –> add all H –> fix charges –> save as mol2 file

2) type *adt* at the system prompt

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

AutoTors

1) Read the ligand molecule

2) Use Autotors to setup the tree and brances

3) Toggle torsion activity

4) Aromatic Carbons C–> A

5) Non–Polar hydrogens (merge or restore)

AutoGpf (grid parameter file)

1) Read the macromolecule

2) If the solvation parameters are not added ADT queries whether to add them, if you say yes, it converts the mol2 file to pdbqs file

3) Set Map types

4) Set grid maps

5) Write GPF or edit GPF

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

AutoDpf (docking parameter file)

1) Read the macromolecule

2) Read the ligand

3) Select Docking algorithm

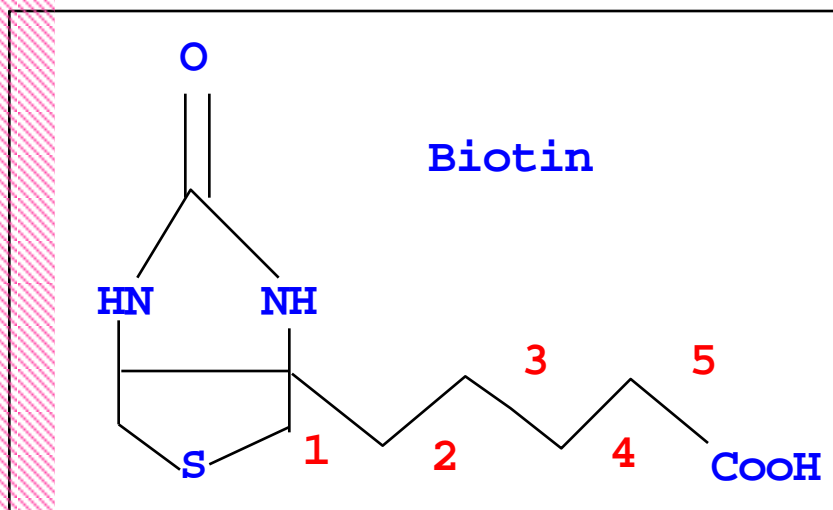**4) Set docking run parameters**

5) write or edit DPF file

Start

1) Start AutoGrid

2) Start AutoDock

3) Options to submit jobs in other SGIs

4) Cancel the jobs

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Test Case

Streptavidin/Biotin (1stp) (2.6 Ang Resol)

Weber et al 1989

Biotin

Autogrid:

Number of points:

116, 104, 124 (x,y,z),

117 X 104 X 124 = 1508832

Grid Spacing = 0.375 Ang.

Map types C,H,N,S,O & e

CPU Time: 7.53s AutoGrid

Octane SGI workstation

Docking: LGA-LS, Population Size = 50;

Elitism = 1, Cross-over rate = 0.8

rate of mutation = 0.02, GA_num_evals 250000

| Lowest Docked Energy Mean	Mean Docked En ergy	Number in Clust er	Ref. RMSD	Est Free Ene. of Binding
–10.78	–10.46	6	0.39	9.03

CPU Time: 3m 25.13s

Octane SGI workstation

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov

Streptavidin/Biotin complex (1stp)

P.C. Weber, D.H. Ohlendorf, J.J. Mendolowski, and F.R. Salemme (1992)

S.Ravichandran, Ph.D., ABCC, NCI–Frederick, MD; sravi@ncifcrf.gov
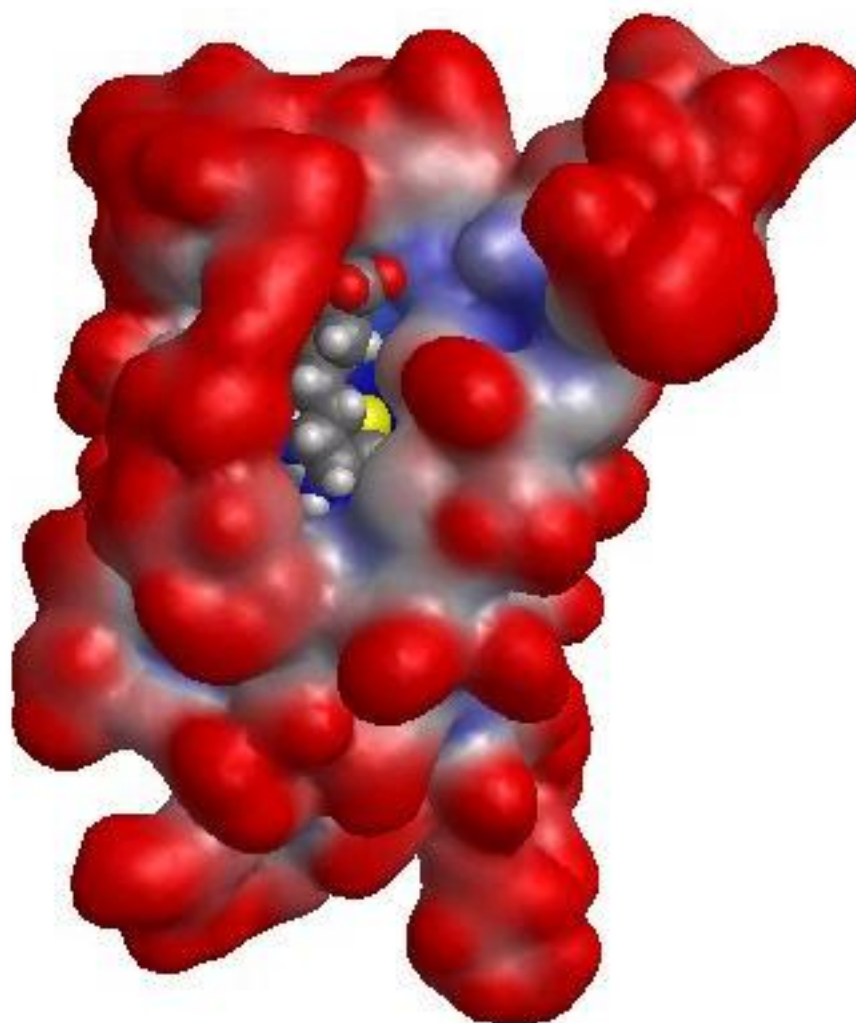
Bound Conformation   **Violet**
Lowest Energy Conformation   **Red**

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov

Things to know:

Not possible to find whether a ligand will bind with millimolar, micromolar or nano-molar binding constants

Genetic Algorithm methods are non-deterministic (sucessive runs may not produce the same answer).
Also there is no guarantee that the solution identified by GA will be the best solution.

Useful Links

http://w3.to/autodock

http://nciiris.ncifcrf.gov/~ravichas/docking

Happy Docki

S.Ravichandran, Ph.D., ABCC, NCI-Frederick, MD; sravi@ncifcrf.gov